DOCUMENT RESUME

ED 107 229 IR 001 968

AUTHOR Pearson, Charls; Slamecka, Vladimir
TITLE Semiotic Foundations of Information Science. Progress Report Number 1.
INSTITUTION Georgia Inst. of Tech., Atlanta. School of Information and Computer Science.
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE Mar 75
NOTE 14p.; Internal Research Memorandum

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Information Science; Information Scientists; *Information Theory; *Language Patterns; Linguistics; Models; *Research; Research Needs; Research Projects; *Semiotics; State of the Art Reviews; Symbolic Language; Syntax
IDENTIFIERS *Semiosis

ABSTRACT

Research on the role of semiosis--a process in which something functions as a sign to an organism--in information processes is described. The object of the study is the structure of various types of signs, and a determination of the relationship between sign structure and information properties. The two goals focused on during the period for which the report was written include: (1) the determination of the internal structure of various sign categories and the relationship between internal and external structure, and (2) formal explication of information including its relationship to semiotic processes and a determination of the essential dimensions of semiosis. Research findings are reported under six headings: (1) theory of sign structure, (2) syntactic developments of theory, (3) investigations into syntactic shape, (4) investigations into semantic and pragmatic structure, (5) project activities, and (6) miscellaneous reports and findings. (Author/DGC)

ED107229

Internal Research Memorandum

SEMIOTIC FOUNDATIONS OF INFORMATION SCIENCE

NSF GRANT GN-40952

Progress Report No. 1

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

CHARLS PEARSON

and

VLADIMIR SLAMECKA

Principal Investigators

March 1975

School of Information and Computer Science
GEORGIA INSTITUTE OF TECHNOLOGY

Atlanta 30332

IR 001 968

ACKNOWLEDGEMENTS

The work described in this progress report has been supported in part by Grant GN-40952 from the Office of Science Information, National Science Foundation. The project staff members responsible for this report are Francis Coyle, graduate research assistant; Irene Gan, graduate research assistant; Charls Pearson, research associate; Keum Shin, graduate research assistant; Phillip J. Siegmann, associate professor; Vladimir Slamecka, professor; and Pranas Zunde, professor.

The report covers the period of January 1974 to February 1975.

INTRODUCTION

The purpose of the present study is to gain a fundamental understanding of the role semiosis plays in information processes. The object of the study is the structure of various types of signs, and a determination of the relationship between sign structure and information properties.

The study started with four specific goals: 1) the determination of the internal structure of various sign categories and the relationship between internal and external structure; 2) formal explication of information including its relationship to semiotic processes and a determination of the essential dimensions of semiosis; 3) a systematic listing of all information measures that have appeared in the literature, along with the known properties of each, classified according to their dimensionality and semiotic interrelationships; and 4) the development of a "kind of directory of results, methods, and key questions in this area which can orient researchers, students, practitioners to facilitate the synthesis and evaluation of research."

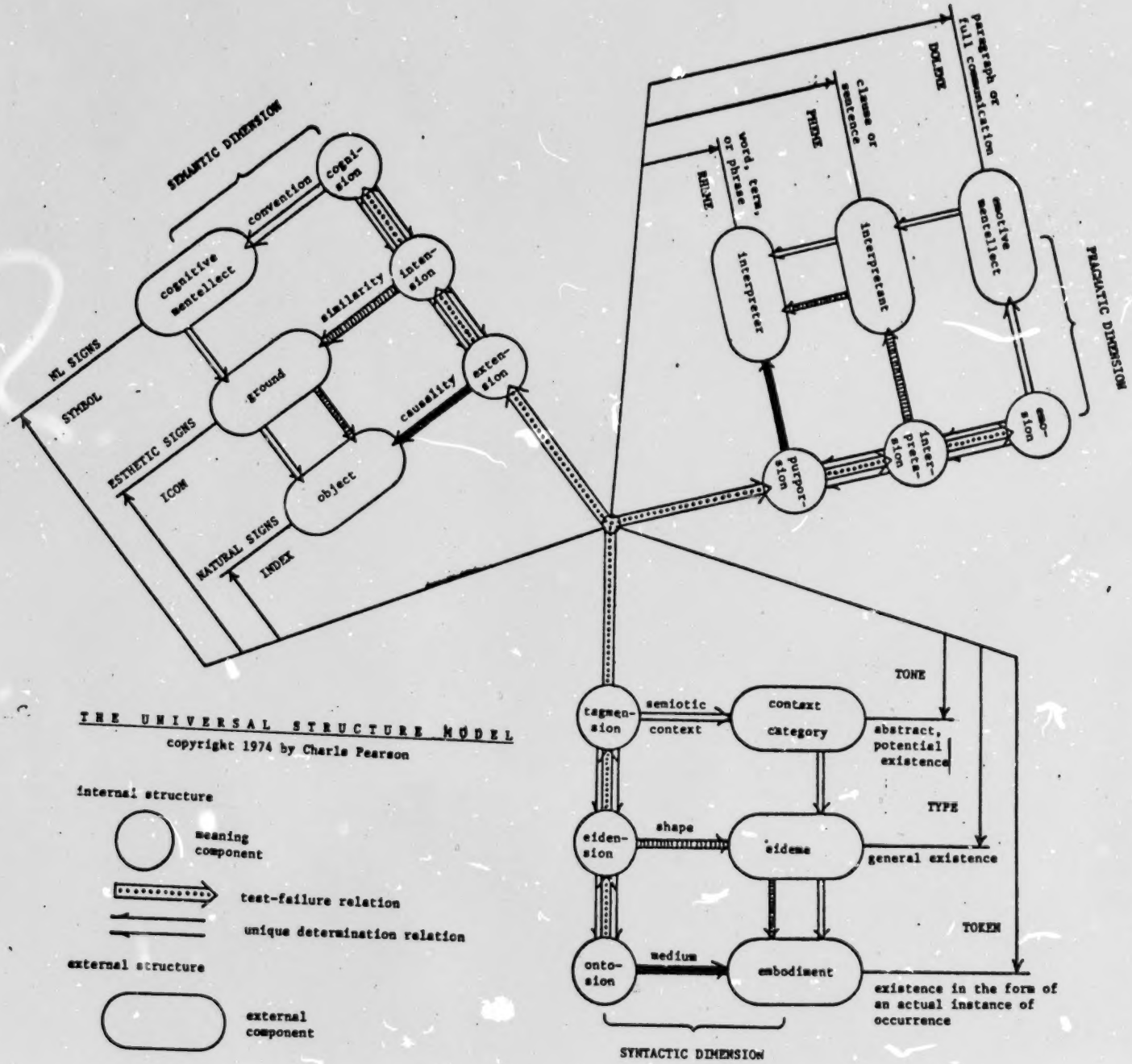
Because of the logical priority of the first two goals over the other two, research during the first year of the contract has concentrated on these. Our progress will be reported under the following six headings 1) Theory of Sign Structure; 2) Syntactic Developments of Theory; 3) Investigations Into Syntactic Shape; 4) Investigations Into Semantic and Pragmatic Structure; 5) Project Activities; and 6) Miscellaneous Reports and Projects.

I. THEORY OF SIGN STRUCTURE

Our purpose in developing a theory of sign structure is to have a tool for explicating the nature of information measurement and its relationship to semiotic processes and for classifying information measures according to their semiotic dimensionality and interrelationships. A theory of sign structure useful for these purposes has evolved gradually over the past two years and its broad outlines have now become quite clear. The theory is called the "Universal Structure Model" and may be summarized by the diagram in fig. 1.

The model is universal in the sense that the structure of all categories of signs is displayed in this one model. For instance an index has only the first, or lowest, level of semantic structure which includes the object and extension of the sign; an icon has the first two levels of semantic structure which include the ground and intension of the sign in addition to the first level structure; and a symbol has all three levels of semantic structure including a cognitive mentellect and cognision.

This theory is an outgrowth of the author's dissertation research into the structure of the symbolic rheme reported in [17]. In that work the meaning of a sign is identified with its internal structure. A separate report on the various senses of the word 'meaning' found in a survey of 20th century literature on meaning and semiotics is being prepared for publication [18]. Another report [19] which explicates the distinction between internal and external structure of signs, explicates the number of sign components, and motivates the identification of meaning with internal components is also in preparation. Some preliminary remarks concerning this theory were published in [20 p4-7].



THE UNIVERSAL STRUCTURE MODEL

copyright 1974 by Charle Pearson

internal structure

meaning component

test-failure relation

unique determination relation

external structure

external component

fig. 1

II. SYNTACTIC DEVELOPMENTS OF THEORY

The universal structure model predicts three levels of syntactic structure: ontotic, eidontic, and tagmatic. The ontotic level contains the medium and ontosion of the sign and is related to the embodiment and other token characteristics. The eidontic level contains the shape and eidension of the sign and is related to the type characteristics. The tagmatic level contains the semiotic context and tagmension of the sign and is related to the context category and other tone characteristics.

In the syntactics of natural language these levels may be identified with phonetics, morphophonemics, and tagmatics, altho this identification has not been explicated yet. Instead, efforts have been concentrated on using these predictions to explicate the statistical theory of syntactical communication. Progress has been quite rapid and it appears that this is the most natural explication for this theory. There is one communication component for each syntactic level. The component for the tagmatic level has been called "information source" [2, chap. 6], the eidontic component is called "encoder", or "decoder", and the ontotic component is the channel. The semiotic properties associated with tone, type, and token phenomena and especially such observed regularities as the rank-frequency law of Zipf and Estoup may be used to understand the communication processes associated with each component. So called "Information Theory" (the Shannon calculus) is seen as the mathematical idealization of certain relations which hold thruout the syntactic dimension (as well as other --- nonsemiotic relations). This approach has been incorporated into the class notes for a senior level course on communication processes [14] and makes these processes quite easy to explain. A textbook incorporating the universal structure model approach is in preparation [23].

III. INVESTIGATIONS INTO SYNTACTIC SHAPE

Original investigations into the nature of syntactic theory have been concentrated at the eidontic level with the semiotic concept of shape being of prime interest. In many kinds of signs, shape is primarily concerned with length and pattern. In 1965 Kolmogorov proposed a measure of shape that is mainly a measure of the pattern. This has been called "algorithmic information" and "complexity", [6]. A historical and tutoral paper was prepared giving a semiotic analysis of this measure as a background for a future explication of the concept of complexity [22]. This is also discussed in [20, p7].

An instrument, called an 'eidontic deviometer' has been developed to measure the deviation of the shape of a natural language sign from a hypothetical norm, or average shape, of a sign in a given natural language. Measurements on artificial word forms using this instrument are both reliable and precise. A paper presenting this development is in preparation [21]. Miller, Bruner, and Postman [8] have shown that the interpretation of signs is affected by their shape. The eidontic deviometer should enable a precise measurement of this phenomena and thus lead to a better understanding of the interpretation process.

From a count of 5½ million letters in a corpus of standard American (the Brown Corpus) a table of polygram frequencies has been prepared. Previous tables for American English available to the public were at least half a century old and their generality is at best suspect, having most likely come from counts of military documents. Several later and more general counts are available within the CIA and the Department of Defense but their access requires a security clearance and a 'need to know". These

tables are useful for the generation of artificial word forms and the study of redundancy in natural language. A paper presenting these tables is in preparation [16]. During the analysis of this count data, a rank-frequency regularity was discovered among the letters. However, unlike the rank-frequency law of Zipf and Estoup for words which is log - log in nature, the regularity for letters is log - linear in nature. A preliminary literature search shows no previous mention of this regularity. All available data for other alphabets and phonemic systems was analyzed and the relationship holds in every case. Much more analysis is required before a paper can be written, but it is hoped that we have discovered a universal relation for the shape elements of a system of discrete signs.

It should be mentioned that Garner's work [4] which is analyzed in a review which is discussed in section V [15] is also pertinent to a study of shape in semiosis.

IV. INVESTIGATIONS INTO SEMANTIC AND PRAGMATIC STRUCTURE

Another area of original investigation which has just begun concerns the semantic and pragmatic structure of signs. Altho the universal structure model stems from research into natural language, if it has any bearing on reality at all, this same structure should show up in other disciplines which study sign processes such as philosophy and psychology.

A preliminary argument has been developed which shows the usefulness of the universal structure model for unraveling philosophical problems. G. E. Moore, a British philosopher of the early twentieth century developed a paradox which has come to be called Moore's paradox of analysis and goes something like this. If the analysis of the meaning of a word has the same

meaning, it is trivial; but if it has a different meaning, then it is wrong. Now, as Moore well knew, philosophers very often make correct and non-trivial analyses, but he was never able to develop a theory of analysis which overcame his own paradox. Other philosophers have tried with varying amounts of success. The problem has never been solved completely. The most popular approach is to say that the problem lies in the formulation of the paradox which assumes that meaning is either a single or wholistic kind of thing which is either completely the same or else altogether different. Frege and Carnap assumed that the meaning of signs has two components, but their assumptions were for entirely different purposes. Carnap was able to delineate the character of scientific analysis fairly well with his "extension" and "intension" but he was never able to handle philosophic analysis. Moore himself said he thought philosophic analysis required something like determining the same objects by the same properties but understanding or cognizing this determination in a different way. Now by looking at our semantic structure and realizing that cognision uniquely determines intension which in turn uniquely determines extension while a difference in extension ensures that two terms will have a difference in intension which in turn ensures a difference in cognision, we may state the solution as: Scientific analysis requires an identical extension with a difference in intension, while philosophic analysis requires an identical intension with a difference in cognision.

An area of psycholinquistics that is eminently suited for investigating sign structure is aphasia. Many attempts have been made to classify various types of aphasic disorders or to establish a taxonomy of aphasia without much success. An attempt has just recently begun to classify aphasic problems by the sign component with which the impairment is associated and the degree of impairment. It is thought this same approach would be successful with pharmocological studies.

One final area we have begun to explore concerns cognitive representation. Kintsch has reported three aspects of cognitive memory which he calls "sensory", "short term", and "long term". Bruner has reported several modes of representation, or coding, including "iconic" and "symbolic". He has studied the sequence in which these capabilities develop in children and the rate at which signs can be processed using thse various modes of representation. It would appear as if there were just one form of coding associated with each aspect of cognitive memory. However, the experimental data is not clearcut on this question because of confounding effects.

An experimental program is being designed to critically isolate each memory aspect and the mode of representation that is associated with it, using an interference effect discovered by Siegmann. The advantage of achieving an answer to this question is to allow quantitative measurements of psychology to be used in future investigations of semantic structure. For instance, memory span times, processing rates, and age of development are all quantitative measurements, and all run in the sequence: index, icon, symbol.

V. ACTIVITIES

One of the investigators (C.P.) organized and chaired two technical sessions on the semiotic foundations of information science at the 1974 annual meeting of the American Society for Information Science (ASIS) in Atlanta. Six papers on various semiotic aspects of communication were presented at sessions number 6 and 8, [1, 3, 5, 7, 9, 24]. At the first of these sessions, an introduction was given outlining the work on semiotics currently underway at Georgia Tech [10].

At this same meeting a special seminar was held to present the universal structure model, [11]. At the Tampa meeting of the Southeastern Conference on Linguistics in May, 1974 (SECOL XI) a paper was delivered outlining the semiotics lab being developed at Georgia Tech under the Auspices of this project [12].

VI. MISCELLANEOUS REPORTS AND PROJECTS

A semiotics lab has been developed which supports several courses in information science. The initial development was reported in [20, p3-4]. A lab manual has been written for the use of the students in this lab [13].

In the entry for information science in the Encyclopedia of Computer Sciences, [25] the authors concluded that semiotics is at the core of information science as a science.

W. P. Garner's latest book on information and structure was reviewed for the ACM's Computing Reviews [4, 15]. This is a report on ten years of research into the nature of medium and shape in the process of sign perception. The review concentrates on an analysis of the semiotic concepts involved.

A textbook which uses semiotic concepts to explicate the measurement of information is being developed [26].

A preliminary progress report on this project appeared in [20].

REFERENCES

1. Abilock, J.; and Yovits, M. C. "Semiotic Models of Information Flow Leading to a Quantitative Measure of Information". Ohio State University.

2. Ash, R. Information Theory. Interscience, 1965.

3. Budd, R. "Semiotic Aspects of Human Communication". Rutgers University.

4. Garner, W. P. The Processing of Information and Structure. Wiley & Sons, 1974.

5. Gough, J., Jr.; and Valach, M. "Some Aspects of Semiotics Which are Fundamental for Information Science." Georgia Institute of Technology.

6. Kolmogorov, A. N. "Three Approaches to the Quantitative Definition of Information". Problems of Information Transmission, 1(1965), p1-7.

7. Ladany, S. P. "Valid Data for Design of Hebrew Language Information Processing Equipment". Columbia University.

8. Miller, G. A., Bruner, J. S., and Postman, L. "Familiarity of Letter Sequences and Tachistoscopic Identification." J. Gen. Psychol., 50(1954), p129-139.

9. O'Hara, R. "Semiotics of the Film". University of South Florida.

10. Pearson, C. R. "Semiotic Research at Georgia Tech - 1974". Introductory Remarks at Technical Session 6, 1974 Annual Meeting of American Society for Information Science.

11. Pearson, C. R. "Peirce's Theory of Signs and the Universal Structure Model". A two-hour seminar presented at the 1974 Annual Meeting of the American Society for Information Science.

12. Pearson, C. R. "Georgia Tech's Semiotics Lab". Paper presented to SECOL XI at University of South Florida, Tampa, Fla. May 3-4, 1974.

13. Pearson, C. R. "Lab Manual for Semiotics". School of Information and Computer Science, Georgia Institute of Technology. Atlanta, Georgia, 1975.

14. Pearson, C. R. "Communication Processes": unpublished course notes for ICS-4120.

15. Pearson, C. R. "Review of Information and Structure by Garner, W. P." In ACM's Computing Reviews, to appear.

16. Pearson, C. R. "A Table of Polygram Frequencies". In preparation.

17. Pearson, C. R. The Concept of Meaning in Natural Language. Ph.D. Thesis; Georgia Institute of Technology, 1975. (In preparation)

18. Pearson, C. R. "The Senses of 'Meaning'." In preparation.

19. Pearson, C. R. "The Components of Signs". In preparation.

20. Pearson, C. R.; and Slamecka, V. "Semiotic Foundations of Information Science". In Research 1973/1974 Annual Progress Report, School of Information and Computer Science, Georgia Institute of Technology; Atlanta, Georgia; p3-7.

21. Pearson, C. R.; Smith, L.; and Gan, I. "The Eidontic Deviometer: An Instrument for Measuring Shape Deviance in Natural Language". In preparation.

22. Pearson, C. R.; and Zunde, P. "The Kolmogorov Potential as a Measure of Algorithmic Information". Submitted to Computing Surveys, 1975.

23. Pearson, C. R.; and Zunde, P. A Semiotic Approach to the Statistical Theory of Syntactic Communication. Textbook in preparation.

24. Rubin, B. "Semiotic Aspects of Personal and Social Communication" Rutgers University.

25. Slamecka, V.; and Pearson, C. R. "Information Science". In A. Ralston, (Ed.) Encyclopedia of Computer Science (Petrocelli Books, 1975), In press.

26. Zunde, P.; and Pearson, C. R. Textbook on information measurement. In preparation.